A High-Performance Software Framework and Interoperable Applications for the Rapid Advancement of Earth System Science

Part III: Data Assimilation Applications for the Earth System Modeling Framework (ESMF)

Principal Investigator:

Arlindo da Silva
Data Assimilation Office, NASA/GSFC, Code 910.3*

Co-Investigators:

Stephen Lord, John Derber, Mark Iredell
National Centers for Environment Prediction, NOAA

Michele Reinecker, Christian Keppenne
Ice and Ocean Branch†, Code 971

John Marshall
Massachusetts Institute of Meteorology

Byron Boville, Cecelia DeLuca
National Center for Atmospheric Research

Peter Lyster
Data Assimilation Office and University of Maryland, College Park

*Laboratory for Atmospheres, NASA Goddard Space Flight Center
†Laboratory for Hydrospheric Processes, NASA Goddard Space Flight Center
Greenbelt, Maryland 20771 USA

November 14, 2000

Contents

Abstract

This is one out of three closely coordinated proposals written in response to the NASA HPCC Earth and Space Sciences (ESS) Project Cooperative Agreement Notice ”Increasing Interoperability and Performance of Grand Challenge Applications in the Earth, Space, Life and Microgravity Sciences” (CAN-00-OES-01). These proposals are aimed at establishing a landmark collaboration between NASA, NOAA, NSF, DOE and the university community to develop a community Earth System Modeling Framework (ESMF).

In Part I we propose the construction of an Earth System Modeling Framework (ESMF) to provide high-performance, extensible and interoperable codes for weather prediction, climate simulations, and data assimilation. The multi-institutional and cross-disciplinary interactions among the broad community involved in the proposed work, together with the greater ease in composing robust and exchangeable components that the framework will provide, will result in a new generation of diverse and performance portable earth science applications.

The framework to be developed in Part I will consist of component coupling services, a set of data constructs that support operations on a variety of data grids and decompositions, and a portable, optimized set of low-level utilities. The linked Part II proposal, addresses deploying community atmosphere, ocean, land and cryosphere models under the framework. In this Part III we will utilize framework services to build atmospheric and oceanic data assimilation applications, namely: 1) atmospheric 3D-VAR data assimilation systems involving components from NCEP, DAO and NCAR; 2) oceanic Ensemble Kalman Filter data assimilation system using NSIPP components; and 3) an infra-structure for building a 4D-VAR oceanic data assimilation system using MIT components. These systems will be built using the foundation classes developed in Part I, extending them when necessary to address specific needs of operational and research data assimilation algorithms.

Key words: Framework, parallel computing, earth system, data assimilation, weather prediction, climate modeling

(I) Science, Technology and Management

1 Introduction

This is the third of three linked proposals to construct an Earth System Modeling Framework (ESMF) that will allow models to interoperate and share components, improve their portability and ease of use, enable them to exploit rapidly evolving computer technology, and insulate them from changes in technology. The ESMF will support codes for weather prediction, climate simulations, and data assimilation. Part I deals with the core development work of the framework. Parts II deals with coupled model applications, while this Part III deals with the specifics of data assimilation applications.

Earth system models are increasingly large and complex endeavours for which new scientific demands have arisen, such as understanding the range of results among different Earth system models, and constraining these models with data. The current volatile high performance computing environment poses a distinct technical challenge for development of applications. The transition from low-parallelism, shared-memory, vector computers to massively-parallel, distributed-memory, cache-based computers is especially challenging for data assimilation given the intensive I/O nature of the application, the unstructured nature of the various data streams, and the global nature of the inverse/minimization problem.

The framework will consist of services for coupling and exchanging data between component models and model subcomponents, a hierarchy of data structures that support operations on a variety of data grids and decompositions, and a portable, optimized set of low-level utilities. A unique feature of our design is an integrated approach to data assimilation for earth systems modeling. General purpose utility routines will include a standard interface to a variety of memory management and communication protocols for scalable architectures, I/O formats and packages, performance profiling, time management, and error and signal handling. The framework will be designed in an object-oriented, layered manner to isolate machine dependencies, and offer an application programming interface natural for the Earth sciences. The data constructs and low-level utilities will be used by the coupling superstructure and may also be used separately to compose scientific applications. Thus our work will promote software reuse as well as i nteroperability.

The principal investigators on the three proposals will initially closely coordinate the definition of an *open standard* covering the application domains that the ESMF will target. This effort has already begun. The implementation of the standard under Part I will then get underway, accompanied by efforts on the part of the participants in Parts II and III to undertake rapid prototyping of key elements of the standard, perhaps using some existing frameworks that partially fulfill the requirements. When the ESMF itself is implemented, the participants on Parts II and III will undertake exhaustive testing and performance studies based on their own applications, and disseminate their results to the wider community.

For this Part III proposal, the leading national centers for operational atmospheric data assimiliation and two of the leading groups for ocean data assimilation for climate applications will collaborate to enhance development and interoperability of assimilation systems for any model designed under the ESMF. The data assimilation applications we will implement in ESMF are atmospheric 3D-VAR assimilation systems, sequential ocean data assimilation systems, and the infrastructure to support 4D-VAR ocean data assimilation. The particpants are NOAA's National Centers for Environmental Prediction (NCEP), focussed on operational numerical weather prediction, NASA's Data Assimilation Office (DAO), focused on both real-time and retrospective analyses of atmospheric satellite observations, NASA's Seasonal-to-Interannual Prediction Project (NSIPP), focussed on assimilation for ocean initialization in coupled ocean-atmosphere-land surface forecasts, and the Massachusetts Institute of Technology (MIT), a major participant in the "Estimation of the

Circulation and Climate of the Ocean" (ECCO) Consortium, focussed on dynamically consistent estimates of the ocean state for climate analyses.

The collaboration between NASA's DAO and NOAA's NCEP being proposed is consistent with the goals of the recent white paper *A NASA and NOAA Plan to Maximize the Utilization of Satellite Data to Improve Whether Forecasts* (Franco Einaudi, *personal communication*). This white paper outlines the required steps to enhance and solidify the synergism between the two agencies, with the goal of better utilizing satellite data for both operational and research purposes. In particular, this white paper calls for a common model and data assimilation infrastruture, as means to accelerate the use of data from advanced satellite sensors.

2 Scientific Objectives and Rationale

2.1 Rationale for the ESMF

The value of interoperable codes for Earth system modeling and data assimilation has become increasingly apparent. Forecast requirements for both weather and climate are becoming more stringent, and data assimilation is becoming a crucial component for both prediction and analysis of the climate record. Data assimilation has its own large and independent community, and its tools are not, in general, accessible across the climate modeling community. Given the complexity of data assimilation systems, much is to be gained from interoperable code. Despite the advantages of code interoperability, very little progress has been made toward this goal. Groups share forward models, e.g., radiative transfer codes such as OPTRAN, but fail to share more high level components such as on-line quality control systems and analysis equation solvers.

The difficulty of the assimilation problem increases with the complexity of the model and the number and different types of observations to be assimilated. The more complex the systems the harder it will be to require interoperability of components. At the same time, it is precisely for the more complex systems — such as coupled climate models and data assimilation systems — that interoperability is most important.

To further complicate the task before us, the requirements for much more sophisticated assimilation methods have coincided with major changes in computer architecture. The need to utilize distributed memory architectures, or even more daunting, to mix in the same computer and the same application distributed and shared memory programming models, has placed a huge burden on the already strained software development efforts of most groups doing data assimilation. It has also led to the realization that the old approach, in which each center develops its own solutions, is not just preventing the interchange of scientific codes, but is simply becoming unaffordable. A common *reusable and interoperable* solution to these problems must be developed. This is the rationale for an ESMF.

2.2 Objectives of the ESMF

The problems and the needs are clear, but it is still important that we present a rigorous case that the development of an ESMF is the solution. A framework will involve political and administrative overheads and will, inescapably, require some compromises in performance and flexibility from all who comform to it. If the benefits do not outweigh these costs, the project is rightly doomed to fail. A particular challenge to the ESMF is that applications in the operational environment with its strict requirements on turnaround cannot afford a marked performance degradation merely to support interoperability. On the positive side, the distinct benefits to be gained from published standards in coding and common interfaces to access data streams should be a strong enabling factor in the transition of new developments from the external community to the operational groups.

The specific objectives of the ESMF are to provide the following benefits:

- **Facilitate the exchange of scientific codes (interoperability)** so that researchers may more easily take advantage of the wealth of resources that are available in the US in smaller-scale, process modeling and to more easily share experience among diverse large-scale modeling efforts.
- **Promote the reuse of standard, non-scientific software**, the development of which now accounts for a substantial fraction of the software development budgets of large groups. Any center developing or maintaing a large system for NWP, climate or seasonal prediction, data assimilation, or basic research will have to solve very similar software engineering and routine computational problems.
- **Focus community resources** to deal with architectural changes and the lack of quality commodity middleware. The non-scientific parts of the codes that would be dealt with in a common framework are also the most sensitive to architectural changes and "middleware" quality.
- **Present the computer industry with a unified, well defined and well documented task** for them to address in their software design. As noted already, the scientific community's influence with the industry is much diminished, but it will be even smaller if it is exercised separately by five or six centers.
- **Share the overhead costs** of the housekeeping aspects of software development: documentation and configuration management. These are the efforts that are most easily neglected when corners have to be cut.
- **Provide institutional continuity** to model development efforts. Most US modeling efforts are necessarily tied to only a few individuals, and centers are hard-pressed to maintain continuity that transcends them. The competitive job market we are in will result in shorter tenure for programmers. At the same time, the increasing complexity of both our systems and the technology will produce more reliance on software specialists — and less on scientists — to maintain these aspects of the systems. Both of these factors will contribute to a more unstable workforce and much greater difficulty in maintaining "institutional continuity." A framework can helps us do this by having a much larger institution to support it — the whole community. Also by having codes depend on a common, well-known framework, it will be easier to find and train new people to continue a line of development.

2.3 The ESMF Community and NASA's Role

The diversity of applications supported by the ESMF is what is required to support the national Earth science objectives and more specifically to support NASA's mission as outlined in the Earth Science Strategic Enterprise Plan 1998-2002 cited in the Announcement. NASA's mission in the Earth sciences focuses on data: its acquisition, dissimination, analysis, interpretation, and use to improve our knowledge of the Earth system and our ability to predict it behavior. An ESMF would make major contributions to NASA's activities in the last three of these.

Today, the analysis of geophysical data relies heavily on detailed and comprehensive numerical models of all components of the Earth system and on sophisticated data assimilation algorithms capable of synthesizing myriad observations into a physically consistent picture. Further use of this information to predict the state of the system days or months in advance, or how it will respond to external perturbations years or decades from now also requires comprehensive models of the atmosphere, ocean and other components.

2.4 Scientific Benefits

In this section we describe ways in which the ESMF will increase scientific productivity and encourage new research in a range of Earth science domains. These include climate modeling and general circulation modeling (focus of Part I), numerical weather prediction, and data assimilation (focus of this Part III). We include some related scientific domains into which we believe the benefits of the ESMF could be readily extended.

2.4.1 Numerical Weather Prediction

Operational forecasting centers have strict requirements for performance and robustness. By specifically optimizing parts of the ESMF, some of the hand-tuning optimization burden at operational centers can be shared with the rest of the modeling commmunity. Also, sharing utility code for error and signal handling, and having the whole community develop and test code, will contribute to the robustness of code overall.

The forecasting codes we will use as a testbed for ESMF are 1) the NCEP atmospheric global forecast code, and 2) DAO's Finite-volume forecasting code which is part of NCAR/NASA-DAO CCSM [?] effort. NCEP's system is a key component of the Global Data Assimilation System (GDAS) that provides that backbone of all numerical weather prediction at NCEP. It also is used to make the 4 times per day 120- hour Aviation forecast, the daily 384-hour Medium Range forecast and the 22 per day 384-hour Ensemble forecasts at NCEP. The DAO system is run operationally in support of NASA/EOS intrument teams as well NASA missions and field campaigns.

2.4.2 Data Assimilation Applications

Data assimilation techniques have been developed and used for many years to provide initial conditions for numerical weather prediction. In more recent years, longer term, climate forecasts have become feasible and desirable, as have model-data syntheses of the myriad of observations of different types and accuracies into a physically consistent picture of the climate record. NASA maintains efforts in all of these areas. In addition, collaborations between NASA and NOAA are focussed toward development of next-generation operational components, to take advantage of work on algorithmic improvements and the incorporation of new satellite data types into operational systems. Technology transition is a significant challenge when it is at the software level, from research and development to the rigorous, stable standards of the operational environment. Thus, work at Goddard's Data Assimilation Office (DAO) and NASA's Seasonal-to-Interannual Prediction Project (NSIPP), and NOAA's National Centers for Environmental Prediction (NCEP), participants in this proposal, would benefit immediately from the development of the ESMF. Ocean products are only now emerging into the quasi-operational arena, so the inclusion of ocean state estimation tools within the same framework as atmospheric state estimation tools will be beneficial to speed development, portability, and interoperability.

At least as complex as the atmospheric/ocean models used, is the analysis software that comprises the core of data assimilation algorithms. As in the modelling community, these systems have been largely developed in isolation, making it difficult for groups to take advantage of development taking place outside their home institutions. The interoperability afforded by ESMF will enable the sharing of high level components such as on-line quality control systems, minimization algorithms and management of observational databases.

The data assimilation applications we will implement in ESMF are diverse and span a variety of dynamical models and assimilation algorithms. Two different global 3D-VAR atmospheric data assimilation systems used operationally at DAO and NCEP will be implemented under ESMF, affording a great degree of interoperability between the two centers. DAO's system will be based on the Physical-space Statistical Analysis System (PSAS, Guo and da Silva 1997, Cohn et al. 1998), while NCEP's system will be based on the Spectral Statistical Interpolation algorithm (SSI, Parish and Derber 1992). Two other systems using NASA/GSFC-NSIPP component models, optimal interpolation and the Ensemble Kalman filter, focus on assimilation for ocean initialization in coupled ocean-atmosphere-land surface forecasts. A complementary approach is utilized for dynamically consistent estimates of the ocean state for climate analyses by the Massachusetts Institute of Technology, one of the major participants in the "Estimation of the Circulation and Climate of the Ocean" (ECCO) Consortium. This effort will develop the infrastructure under ESMF for 4D-VAR ocean data assimilation using the mitGCM and the tangent-linear and adjoint model compiler (TAMC) of Giering and Kaminsky (1997).

2.4.3 Climate modeling and GCMs

The ESMF will provide ready-made solutions to standard climate model issues such as grid representation and transforms, utility software and high-level scheduling. Standard interfaces will help keep the diverse community of developers coordinated.

Climate models are increasingly being used to guide policy decisions. The predictive requirements are becoming more stringent and data assimilation a crucial issue. The demand for interoperability of climate model components has intensified as growing evidence of anthropogenic climate change has focused scrutiny on the capabilities of the current generation of climate models. Without exchangeable model components, it is often difficult to point to a particular component as a clearly identifiable cause of divergent results when one model is compared against another or against the observations. The ESMF will facilitate just this sort of analysis.

The computational demands of coupled climate models necessitate the use of scalable parallel computational platforms. To effectively utilize such platforms codes must be based on flexible data structures and communication libraries that are performance portable. The ESMF will provide this infrastructure.

The most resource-intensive components of climate models are typically oceanic and atmospheric general circulation models (GCMs). Within a GCM physics and dynamics can be treated as separate coupled components themselves. This allows researchers to swap in dynamical cores more easily, thus offering the opportunity to explore more efficient algorithms. Some GSFC-DAO models are already structured this way using the GEMS framework, and plug-compatible dynamical cores for the the atmospheric portion of the NCAR CCSM are currently being developed.

The climate models we will implement using ESMF are the NCAR CCSM [**?**], the NASA/GSFC-NSIPP climate model, the MITgcm [**?**], and the GFDL climate model.

2.4.4 Space Weather

Space weather applications have many of the same features as Earth system applications: coupling of different physical domains, each with different data structures and time scales, a similar basis in fluid dynamics equations, a need for generic utilities, and greed for computational resources. The relation between the two domains is even more direct in efforts such as Michigan's KDI, which is attempting to couple space weather models with Earth system models to study Sun-Earth interactions. We will actively try to coordinate with and engage this community so that we can leverage each others efforts.

2.5 Customers and Delivery of the ESMF

The initial customers of the ESMF are the application groups described in Parts II and III of this proposal set. These include the broad spectrum of university and government researchers who use codes such as the NCAR Community Climate System Model and the mitGCM, operational centers such as NCEP, and data assimilation groups such as NASA/DAO, NSIPP, and ECCO. Other near-term customers for the ESMF include research groups that have been active in the CMIWG effort, such as the Weather Research and Forecast model, and the International Research Institute for Climate Prediction (IRICP), from which letters of endorsement are attached.

In addition, the delivery and promotion of the ESMF will proceed on a number of fronts:

- A substantial number of applications will be entrained in using ESMF through direct relation to our individual Investigator's activities. This includes conversion of other codes at institutions participating in this proposal; codes from collaborators of who will wish to interoperate with our applications and will be exposed to the ESMF through major community events, such as the yearly CCSM workshop; and other groups active in the CMIWG, the impetus for our proposed collaboration. Some of these closely related codes may be introduced as plug-in applications. Both NSIPP and ECCO are participating in the international Global Ocean Data Assimilation Experiment (GODAE). The interactions and collaborations through GODAE will expose the ocean data assimilation community to the ESMF.

- Virtually anyone, internationally, with an interest in Earth system modeling will have access to the ESMF source code and documentation via the web. We will encourage moderated open source development.

- While we will strongly encourage community contributions to the framework, we feel that it is essential for one major center to promote, maintain and moderate continuing development of the ESMF. NCAR has committed to this task.

- A workshop is planned at the end of the proposed work to NASA to introduce the ESMF to a broad community. NCAR's ongoing support role will include the coordination of regular workshops.

- We plan to present our work at conferences and publish our experiences and results in technical and earth science journals

3 Technical Plan

The community collaborating on this proposal has already made significant progress in defining the structure of the ESMF. In June 2000, each participating group agreed to prepare a strawman document describing their application requirements, ESMF scope and architecture, and implementation strategies (see http://www.scd.ucar.edu/css/NASACAN.htm). The initial degree of convergence was high, and a combined strawman was created. Collaborators met at NCAR in August 2000 to review and refine the combined strawman, and work on a more detailed design document for ESMF began. This work is underway. However, we have already converged upon a preliminary design, as outlined here.

3.1 ESMF Functionality

3.1.1 Scope

The scope of the initial ESMF must be extensive enough to offer significant advantages to the Earth science community but modest enough to be completed with the resources offered through this Cooperative Agreement. There are two critical needs of the Earth science community: 1) robust, optimized, non-scientific *infrastructure* libraries with which to build models andmodel subcomponents, to promote code reuse and 2) a *superstructure* for coupling scientific components, to promote code interoperability.

In data assimilation applications, the coupling *superstructure* will perform regridding, interpolation and communication of gridded, distributed data. The data may represent multiple fields or a single field or an observational data stream, may be in the same or different executables, may be in code segments executing serially or concurrently, and may be distributed among nodes and/or partitioned among multiple threads. The interfaces for components and couplers will embody an *open standard*.

The software necessary to support the above capabilities includes *infrastructure* for describing a wide variety of grids and decompositions, and for performing high-level manipulations of fields discretized on those grids, as well as unstructure observational data streams The software for specifying decompositions will interface to a mechanism for performing dynamic load balancing. Operations on grids and fields must implement corresponding methods for the construction of tangent linear and adjoint models for data assimilation applications.

Both the coupling mechanism and application codes use common utility routines. This part of the ESMF *infrastructure* includes communication libraries, sychronization, optimized I/O, performance profiling, time management, and signal and error handling.

3.1.2 Requirements

In addition to achieving the ESMF overall goals (interoperable components, code reuse, simpler code maintenance, etc.) the ESMF code will conform to a set of functional requirements. These include:

Performance portability. Portability and computational efficiency over a wide range of platforms are essential. The framework should not significantly degrade performance of an existing code written without the framework. Optimized performance on scalable architectures for moderate numbers of processors (16-500) is the highest priority.

Flexible usage. The application writer should be able to choose how much or how little of the ESMF to use. For example, data assimilation applications may initially use ESMF only for the data structures and external shell of the inversion procedure and maintain the existing solver software as an external component.

Ease of use. A key principle of design is that every user is also a developer: the key "users" of the framework are "developers" of component models, and the ESMF must simplify, or at least not unduly complicate, their lives.

Extensive grid support. ESMF must be able to couple components that are discretized on: logically rectangular grids (which may be on a physically non-rectilinear metric and a variety of positional stencils, and including "exotic" but logically rectangular grids such the bipolar and cubed-sphere grids); reduced (cut-out or Kurihara) grids; unstructured grids (e.g., land grids, observations); phase space grids (e.g., spectral, Fourier); nested and adaptive grids; and icosahedral grids. In addition we require support for describing masked regions and halo regions.

High performance, extensible, multi-format I/O. The ESMF utilities will support a generic interface for I/O of self-describing data in netCDF, binary, GRIB, BUFR and EOS HDF data formats. Other formats such as the ODF data formats may be added later. There must be support for high-performance parallel I/O.

Multiple language bindings. ESMF utility and coupling software will be usable by applications written in C/C++, F77 and F90.

Other requirements. These include appropriate error and signal handling, runtime configurability, and an efficient, low maintenance implementation (e.g., auto-documentation from code).

3.2 Architecture

The architectural details of the ESMF are summarized briefly below.

3.2.1 Interoperability and reuse

Object-oriented design and design reuse - both standard framework techniques - are the methods we will use to achieve interoperability. By establishing a standard set of methods and coupling interactions, researchers across the country will be able to prepare codes that they can be confident will operate with a large set of others.

We will adopt an object-oriented, layered approach to the ESMF architecture. Object-oriented design enhances code reuse since a class structure encourages well-defined, general purpose, encapsulated code segments that can work in a variety of contexts.

A framework implies that overall design - how classes interoperate - is reused as well as code. In the ESMF design we propose, certain classes within an application will be provided by the framework and others will be supplied by the application developer. The classes provided by the developer must possess a core set of methods in order to interoperate within the framework. For example, an ocean model component of a climate model might need to contain a method that returns a description of its data grid and distribution. This information would be used by a generic data assimilation mechanism to understand how data should be routed between that ocean component and the data inversion (assimilation) component.

3.2.2 Performance portability and ease of use

We will achieve performance portability and ease of use by layering code, through the use of generic interfaces and by designing interfaces that are not burdensome for developers seeking to integrate their applications into ESMF.

Machine-dependent code will be isolated to the lowest level in the framework by wrapping it in a generic interface. Likewise, calls to higher-level communication functions such as transposes are isolated to a layer in the framework so that an application developer does not need to manage the details of distributed data transfers. This makes the application code easier to write and use.

3.2.3 Preliminary design

We have converged on a preliminary design for the architecture of ESMF that we expect will fulfil the requirements of the Earth systems modeling community as laid out above. We describe here the layering strategy, with examples of the kinds of operations native to each layer.

At the bottom are **low-level utilities**, such as communication and memory management primitives, error and signal handling, timing, machine primitives, and basic I/O operations. These are often machine-dependent and may be coded procedurally for efficiency.

The second layer is a set of **parallel utility classes**. This layer includes a retrievable description of a *machine model* and a layout class that describes the portion of a machine over which a data object is distributed, and associates a simple topology with it.

Classes representing **distributed grids** are in the third layer. Distributed grids contain a layout and a grid specification that describes the grid coordinates and connectivity. A substantial portion of the ESMF resides in this layer. The methods here include *index-space* and *physical-space* methods.

Fields are in the fourth layer. Fields contain a distributed grid, and a field specification that describes attributes related to the physical field ("metadata"). We will also support a *field bundle* class, for fields that are discretized on the same distributed grid. High-level I/O operations reside in this layer. These will use the metadata information to create comprehensive header information for the self-describing data formats the ESMF will support, and the distributed grid information to create efficient, high-performance, parallel I/O in various modes, including single-threaded, multi-threaded, and distributed I/O.

The **coupling and components** layer includes large scale components, such as atmosphere models and data assimilation (analysis) components, and the classes used to simplify the transfer of data between them. This layer will provide *boundary state vector* objects that comprehensively describe the portion of a component model state that is necessary for coupling, and the operations thereon. This layer also establishes communicators between component models, and the highest-level control module for scheduling them, for serial or concurrent execution.

3.3 Testing and Benchmarking of the Core ESMF

The Core ESMF being developed in Part I will be tested using a synthetic benchmark suite consisting of a set of computationally representative segments from the Earth system codes described in detail in Parts II and III. These will be assembled in close conjunction with the dispersed team of application developers tasked with converting codes to ESMF. The benchmark suite will be baselined using the ESS Testbed, and progress on the ESMF will be measured via code improvement milestones applied to the benchmark suite. The metrics will be time to solution, which should remain the same or decrease as codes adopt the framework; and increased functionality in three areas: a SPMD/MPMD option, a hybrid programming paradigm, and interoperability of C++/Fortran90 components.

This test suite will evolve with the framework and will be maintained as part of the framework. In this way the test suite will provide a rapid way to test and assess the framework as it is developed. Additionally, sites installing the framework and sites wishing to benchmark the framework will be able to use the test suite. These kernel tests will also serve as a possible focus to engage the computational science research and engineering community in the challenge to boost performance and scalability.

In this Part III we will concentrate on validating the framework in the field, by aggressively deploying a set of existing production data assimilation systems.

3.4 Application codes for Data Assimilation

The atmospheric data assimilation applications will utilize the full suite of real-time observations available from NCEP. The ocean data assimilation applications will assimilate subsurface temperature data from expendable bathythermographs (XBTs), which are available at unstructured, time-varying locations, from moored locations (i.e., fixed locations) and surface altimeter data (time-varying swath data) available from GODAE servers.

3.4.1 DAO Atmospheric Data Assimilation System

For integration into ESMF we will concentrated on DAO's next-generation Finite-volume Data Assimilation System (fvDAS) based on NASA-NCAR General Circulation Model and the Physical-space Statistical Analysis System. This system consists of the following main components:

General Circulation Model. The General Circulation Model used in fvDAS is the model jointly developed by the Data Assimilation Office (DAO) and the Climate and Global Dynamics Division (CGDD) at NCAR. This model is based on the *finite-volume dynamical core* developed at DAO (Lin and Rood 1996, Lin and Rood 1997, Lin 1997, Lin and Rood 1998) with physical parameterizations from the NCAR CCM3 (Kiehl et al. 1996). During this project this model will be upgraded to the atmospheric component of NCAR's CCSM.

Quality Control. The Statistical Quality Control (SQC) System is used to screen observational data prior to assimilation (Dee et al 2000). This QC system consists of simple check of the observations against a background field, followed by an adaptive buddy check which adjusts error bounds according to the *flow of the day*.

Analysis System. The Physical-space Statistical Analysis System (PSAS, Guo and da Silva 1997, Cohn et al 1998) is used to combine a first guess from the NASA-NCAR GCM with observational data to provide an updated estimate of the state of the atmosphere.

The NASA-NCAR GCM is a completely new model which which replaces the GEOS GCM used in the GEOS-1/2/3 Data Assimilation systems; see DAO (1996) fora description of the GEOS GCM. A particular configuration of SQC and PSAS are currently implemented in GEOS-Terra system which is run operationally at DAO. The unique finite-volume formulation of the NASA-NCAR GCM, combined with the generality of the observation-space formulation of PSAS, call for a complete redesign of the current GEOS DAS model-analysis interface which has its roots in the *Optimum Interpolation* (OI) algorithm of GEOS-1 DAS (Pfaendtner et al. 1994). The current implementation of fvDAS already includes object oriented concepts and will much benefit from the ESMF being proposed.

3.4.2 NCEP Atmospheric Data Assimilation System

The NCEP atmospheric global forecast code is a key component of the NCEP Global Data Assimilation System that provides that backbone of all numerical weather prediction at NCEP. It also is used to make the 4 times per day 120- hour Aviation forecast, the daily 384-hour Medium Range forecast and the 22 per day 384-hour Ensemble forecasts at NCEP. This hydrostatic sigma coordinate model carries surface pressure, temperature, horizontal winds, moisture, ozone and other tracers as its prognostic variables. The code uses the spectral transform method to compute horizontal derivatives, to solve the semi- implicit Helmholz equation, and to apply subscale horizontal diffusion. The physical grid is used to compute single column physics, including clouds, solar radiation, longwave radiation, gravity wave drag, surface layer exchanges, planetary boundary layer vertical mixing, shallow convection, deep convection, large-scale condensation, and ozone chemistry. The code uses the transpose strategy to distribute data and work across processors. The entire model state is transposed several times every timestep, which for the operational T170 L42 resolution is 450 seconds. The code currently runs on an IBM SP using MPI for communications.

The NCEP atmospheric global analysis code is the second key component of the NCEP Global Data Assimilation System. It also is used to initialize the 4 times per day 120-hour Aviation forecast. This three-dimensional variational code solves a single minimization problem at a specific time given all the received observations within a 6-hour window and their assumed errors along with a forecast guess within the window and its background error covariance. The minimization problem also includes constraints that ensure that the final analysis will be dynamcially balanced as well as fitting the observations and the guess. Many observations such as the satellite radiances require both a guess interpolation and a single column forward radiative transfer model to compute the guess radiances and hence the observation residuals. Therefore, some observations require more computations than others in the analysis. The code performs hundreds of iterations using the conjugate gradient method to perform the minimization. Each iteration contains spectral transforms and so the transpose strategy is again generally used to distribute data and work across processors, but the three-dimensional solver has a particularly high demand on the communications. The code currently runs on an IBM SP using MPI and IBM extensions for communications.

3.4.3 NSIPP Oceanic Data Assimilation System

Two of the leading efforts in ocean data assimilation in a scalable parallel computing environment are those of the NASA Seasonal-to-Interannual Prediction Project (NSIPP), with a focus on assimilation for ocean initialization in coupled ocean-atmosphere-land surface forecasts, and the "Estimation of the Circulation and Climate of the Ocean" (ECCO) Consortium, focused on dynamically consistent estimates of the ocean state for climate analyses. The M.I.T. participants in this proposal are also partners in the ECCO Consortium. In addition to porting to ESMF services, these two ocean assimilation efforts undertake to improve the scalability of codes and take advantage of HPCC platform capabilities to increase problem size or decrease time to solution.

NSIPP 's current operational assimilation is OI with simple prescribed error covariance functions. Under this proposal the OI software will be converted to use ESMF observation classes and the ESMF transformation services to calculate the innovations and to prepare the forecast and observational error covariance matrices prior to inversion. The plan is to transition this ODA system to the Ensemble Kalman Filter (EnKF, e.g., Keppenne and Rienecker, 1999) which is a prognostic calculation of the multivariate forecast error covariances from an ensemble of ocean simulations running as asynchronous objects in a parallel design. The assembly of the covariance matrix will be converted to use the communications and transformation services of the ESMF. The development undertaken within ESMF will ensure the interchangeability of the EnKF with NSIPP's simpler covariance modeling using OI methodology. Commonality with 3DVAR atmospheric assimilation will also be explored.

The NSIPP OGCM is version 4 of the Poseidon quasi-isopycnal reduced gravity model developed by Paul Schopf (e.g., Schopf and Loughe, 1995). This version has been fully ported to generic parallel architectures using a Fortran90 object-oriented design and the GEMS communications framework developed by Max Suarez. The current Pacific configuration, at a resolution of approximately $1/3^\circ \times 1^\circ \times 20$ layers, runs one month of simulation on 64 PEs in approximately 20 minutes on a fully-loaded system. A simple OI algorithm, implemented such that full-domain covariance matrices (using prescribed covariance functions) are assembled in observation space and inverted at analysis time, takes about 30 minutes per month, assimilating every 5 days on 64 PEs. A more sophisticated, multivariate OI algorithm (e.g., Rienecker et al., 1999) updates subsurface temperature, salinity and currents using satellite altimetric observations of sea-surface height. Because of memory limitations, the required multivariate forecast-error covariances are read from a file at analysis time, and a 1-month assimilation cycle takes almost 2 hours on 64 PEs. None of the current assimilation software incorporates observations classes in the same detail that would be included under ESMF.

3.4.4 MIT Oceanic Data Assimilation System

The ECCO ODA system is an alternative methodology – a 4DVAR which requires adjoint transformations to be built into the ESMF. The ocean state estimation system is based on the M.I.T. GCM with the constrained optimization method requiring the tangent-linear and adjoint compiler (TAMC) of Giering and Kaminsky (1997). The initial temperature and salinity fields and the time-varying surface forcing are the chosen controls of the system.

The M.I.T. General Circulation Model (mitGCM) is a versatile tool for simulating both small-scale and planetary scale ocean circulation. The system is designed to run efficiently on a broad range of parallel hardware. The model implementation supports automatic generation of tangent-linear and adjoint forms. This involves the use of an adjoint compiler (ref http://puddle.mit.edu/ ralf/tamc/tam). The mitGCM group will participate in the design, implementation and validation of adjoint forms of framework operators. The adjoint forms of the operators will be compatible with the automatic differentiation strategy employed by the ECCO consortium adjoint data assimilation project (refs http://www.ecco.ucsd.edu, http://ecco.jpl.nasa.gov/).

3.5 Integration of the data assimilation systems into the ESMF

The specific applications envisioned for this proposal will extend the ESMF basic classes for, e.g., specific physical models, observational data and metadata. Each of the prognostic models used in the assimilation applications, as they exist at present, employ functions somewhat analogous to the ESMF elements. Over the course this project, as the core ESMF evolves, the individual modeling systems will transition to the ESMF. The team from this proposal will interact with the ESMF core design team to achieve interopable data assimilation applications with swapable models/3D-VARs. For example, at the "Fields" level, we envision a class powerful enough to represent observational data and metadata, with methods for accessing (e.g., BUFR for atmospheric and oceanic data, or whatever is appropriate for the GODAE servers) observational databases. We also envision a high level interface (through the observation field class) to forward models such as OPTRAN. In addition some applications will transition the full assimilation software to use the ESMF.

Tests will be conducted in two phases. First, the assimilation software will be exercised with an offline driver (i.e., not directly connected with the forward model). Second, the interface directly to the forward model will be tested.

3.5.1 Driver Tests

The initial tests will be conducted by with an offline driver which provides the first guess field from the model, exercises newly developed interfaces with the observational data and calculates the innovation vector (vector of the difference between the observations and the transformed forecast fields in observation space).

3.5.2 Full System Tests

The unit testing described in section 3.5.1 will identify places in the assimilation software where changes are required in order to transform them into ESMF components. In particular some functions that exist within the existing software, such as interpolation to unstructured data locations, will be implemented and tested in the coupling layer. Once it has been demonstrated that these functions are working correctly in the ESMF, the reference models will be changed to use those functions in place of their own native forms. In some codes this will be achieved initially by creating a "shell" routine that maintains the historical interface, other codes will be altered to use the ESMF functions directly. In the full system test, the assimilation increment will also be communicated back to the forward model.

A technical milestone that we will monitor during this phase is the degree of overlap in support code used by the data management and inversion/minimizing software components.

3.6 Interoperability Experiments

The full-blown ESMF-compliant components, derived from the existing codes, will be employed in a series of interoperability "tests". The first tests will be simple experiments involving verifying computational interoperability – exchange of quality-control software, and 'coupling' between different assimilation software and different forward models, even between assimilation software designed separately for atmospheric or for oceanic applications.

3.7 Performance, portability and scalability experiments

A major challenge of framework based systems is avoiding heavy performance overheads. Our architecture allows numerical codes to remain in their native implementation language and to continue to use native code for most numerical operations. This should ensure strong performance will be preserved for codes that are already highly efficient. However, the framework will need to be scalable at both the driver level and at the support layer level. Performance analysis of both parts will be carried out in detail using both driver test codes and full components. These tests will be used to make sure that the implementation of these layers can give satisfactory performance.

3.8 Expertise in Scalable Grand Challenge Applications

Peter: *Could you costumize this section for Part III?*

All participants in this proposal are heavily involved in the development of Grand Challenge Earth science applications for massively parallel systems. Team members have been involved in this work for many years and have had experince with most MPP systems that have been produced over the last 15 years. Currently all participating institutions have highly optimized parallel implementations of at least some of their production codes. Figure **??** shows speed-up curves for a sampling of these applications. These were chosen to highlight the breadth of applications and platforms being used. Shown in the figure are results for both atmospheric (NCEP, NSIPP, CCM) and ocean (POP) models. Two of the models are spectral and two grid point; two are operational codes (NCEP and NSIPP) and two are research codes (CCM and POP). We have chosen to show results on platforms from four of the leading U.S. manufactures: SGI, Cray, IBM, and Compaq. Most of these models, however, run on multiple platforms. Additional results appear in Parts II and III.

3.9 Vendor Support

We have contacted major hardware and software vendors in the scientific computing marketplace to inform them of our proposed community effort and request their participation in benchmarking, design documents and code reviews. We have also requested enhanced technical support. Early input from vendors on data structures and other critical design decisions will help prevent major recoding efforts as we move from prototypes to optimized codes. Participation from vendors in code reviews and benchmarking, together with strong technical support contacts, will help us optimize codes more effectively. [Add vendors who will participate.]

4 Management Plan

4.1 Investigator Team

The Principal Investigator of the proposal is **Arlindo da Silva**, Meteorologist, Data Assimilation Office, NASA/GSFC. He has led the development of the DAO's Finite-volume Data Assimilation System and its transition to parallel architectures.

The Co-Investigators are:

Stephen Lord, Director, Environmental Modeling Center, National Centers for Environmental Prediction;

John Derber, Meteorologist, Environmental Modeling Center, National Centers for Environmental Prediction has designed the NCEP operational atmospheric data assimilation and is one of the co-developers of the ocean data assimilation system used quasi-operationally for seasonal-to-interannual prediction;

Mark Iredell, Meteorologist, Environmental Modeling Center, National Centers for Environmental Prediction;

Michele Rienecker, Oceanographer, NASA Seasonal-to-Interannual Prediction Project; NASA/GSFC, heads NSIPP, and is a co-developer of the NSIPP parallel ocean data assimilation system;

Christian Keppenne, Senior scientist, NASA Seasonal-to-Interannual Prediction Project; General Sciences Corporation, NASA/GSFC, leads the parallel implementation of the NSIPP ODAS and has extensive experience in parallel software design;

John Marshall, Professor of Atmospheric and Oceanic Sciences, Massachusetts Institute of Technology, developed the mitGCM;

Byron Boville, Senior Scientist and Head, Climate Modeling Section, NCAR, was a founder and is a co-chair of the Community Climate System Model (CCSM) project. The CCSM is distributed to and supports hundreds of researchers internationally.

Cecelia DeLuca, Software Engineer, Scientific Computing Division, NCAR, was a co-developer of the STAPL framework for real-time signal processing at MIT Lincoln Laboratory, a software tool currently used in multiple military radar systems.

Our Investigator Team possesses the combination of skills and backgrounds necessary to help guide the design of the ESMF to support Earth system models and data assimilation, to develop the data assimilation applications under the ESMF, and establish it as a standard throughout the climate and weather communities. All of our investigators have experience constructing high-performance codes on parallel platforms.

The data assimilation applications will be based on the existing operational and quasi-operational systems at NCEP, DAO, NSIPP and on the system developed at MIT. The close proximity of the first three will facilitate the coordination of input to the design of the ESMF to support assimilation applications. It also eases coordination of the interoperability tests that are part of the demonstration of the utility of the ESMF. The team members will maintain contact through regular meetings and teleconferences.

4.2 Oversight Teams

As Principal Investigator, Dr. da Silva will serve as the primary contact and administrator of the proposed work. He will negotiate agreements with NASA and among Investigator Team members, and will arrange for disbursement of funds after payment. He will supervise the overall activities of the Investigator Team and help promote the ESMF project to the wider assimilation community. Dr. da Silva will also participate in appropriate *Oversight Teams* for the ESMF core software. These Teams will closely track and guide the design and implementation of the ESMF software. Each Oversight Team will consist of a mix of physical scientists, computer scientists, and software engineers, with some individuals on multiple Teams. The Oversight Teams correspond to different layers of the ESMF software (low-level utilities, fields and grids, coupling), and reflect different interests and expertise. Oversight Teams will remain in close contact with an *Implementation Team* located at NCAR through regular teleconferences and meetings. The Oversight Teams will participate in requirements analysis, design reviews, prototyping and testing over the course of the project.

4.3 Management summary

The key challenge in creating a management plan for the ESMF is to entrain broad expertise in the framework's development and implementation in several applications while ensuring that work can proceed efficiently, and that decisions can be made in an unambiguous manner. We believe that the Investigator Teams and management structure we propose across the three linked proposals will do exactly this.

Our management plan will accomplish the following:

- engage a broad spectrum of the Earth system modeling community in the specification of requirements and the overall design of the framework to maximize expert input and user buy-in;
- utilize groups with more focused interests to oversee the design and implementation of specific framework components in order to achieve timely, informed decisions;
- delegate much of the work of design drafts, prototyping and production coding to a closely integrated, central team of software engineers;
- resolve inconsistencies and differences of opinion throughout the project by allowing final software engineering decisions to be made by the software manager and oversight team leaders.

(II) Software Engineering Plan

In this section we present a software engineering plan for the ESMF, including software team structure and management, a software process that extends from system specification through distribution and maintenance, and tools and techniques to support development, collaboration, and distribution. Issues primarily relating to ESMF design, such as strategies for interoperability and reuse, are described in the Technical Section.

1 Software Teams and Management

1.1 Core ESMF Development at NCAR

An *Implementation Team* will be established at NCAR consisting of five software engineers supervised by a *software engineering manager*. One of the software engineers will be an *integrator* responsible for Team support functions such as configuration management and defect tracking. The Implementation Team will draft design specifications, prototype and implement ESMF components, test and validate the framework, and distribute releases.

Software development will be guided by three partially overlapping *Oversight Teams* focused on different aspects of the framework: utilities, fields and grids, and coupling. The members of the Oversight Teams will include the co-investigators of this proposal, and will consist of appropriate mixes of software engineers, application scientists and computer scientists. Each Oversight Team will designate a lead. Responsibilities of the Oversight Teams will include reviewing software design and tracking implementation progress.

The software engineering manager and integrator will maintain a system view and ensure that development is coordinated. The Oversight Teams and Implementation team will work closely with software engineers with software engineers converting applications to the ESMF, as described in Parts II and III.

1.2 Data Assimilation Application Development

An *Implementation Team* will be established at NASA/GSFC and NCEP consisting of five software engineers supervised by a *software engineering manager*. One of the software engineers will be an *integrator* responsible for Team support functions such as configuration management and defect tracking. The Implementation Team will draft design specifications, prototype and implement ESMF components, test and validate the framework, and distribute releases, in close collaboration of the Core ESMF team at NCAR.

Software development will be guided by two partially overlapping *Oversight Teams* focused on atmospheric and oceanic data assimilation applications. The members of the Oversight Teams will include the co-investigators of this proposal, and will consist of appropriate mixes of software engineers, application scientists and computer scientists. Responsibilities of the Oversight Teams will include reviewing software design and tracking implementation progress.

The software engineering manager and integrator will maintain a system view and ensure that development is coordinated. The Oversight Teams and Implementation team will work closely with software engineers with software engineers converting applications to the ESMF.

2 Software Process

The Implementation Team will follow a structured software process commensurate with CMM Level II [5, 6]. The process will include many of the procedures recommended by the Software

Best Practices Initiative [1], such as software interface specification before implementation. Our documents and reviews will be simplified versions of those described in standard references [3, 7, 8]. We will aim for an effective process free of extraneous overhead.

2.1 Staged Software Development

The major milestones described in Part V are the result of the coordinated completion of many smaller events, each of which has a "completion gate." These gates may be the approval of a document or the verification that a software test has been passed. Table 1 shows the progression of events in ESMF sofware development, and the product and gate associated with each event.

The initial set of events, labelled "ESMF Definition" is focused on specifying the ESMF system and procedures as a whole. The second group of events, "Class Implementation" describes the development steps applied to individual software classes. As classes are completed they will be integrated into an evolving prototype of the ESMF. The final development stage, "Integration and Distribution", involves the integration of classes leading to a software release. The ESMF will have three major software releases corresponding to milestone I,J, and K; smaller releases and demonstrations will be scheduled to insure that the project is on track.

Table 1 ESMF Software Event Progression

Event	*Product*	*Completion Gate*
	ESMF DEFINITION	
Requirements specification Brief document outlines ESMF functional scope and general requirements.	Requirements Document *Prepared by:* all collaborators	Document review *Reviewed by:* all collaborators
Architectural description Shows layering strategy and describes function and interaction of major components. Incorporates Requirements Document.	Architecture Document *Prepared by:* all collaborators	Document review *Reviewed by:* all collaborators
Software process definition Documentation of software procedures for the implementation team. This document will evolve as the project progresses.	Developer's Guide Document *Prepared by:* integrator	Document review *Reviewed by:* software manager
Implementation study Assesses existing software, optimal language, threading strategy and other issues as identified.	Implementation Report *Prepared by:* implementation team with collaborator input	Document review *Reviewed by:* all collaborators
Software implementation and test plan Plan for class implementation and testing based on milestones, class dependencies, and existing code.	Software Implementation and Test Plan *Prepared by:* software manager	Plan review *Reviewed by:* all collaborators
	CLASS IMPLEMENTATION	
Class design Includes requirements, function, and interface specification. Requirements may be reviewed first for complex classes. Groups of classes may be reviewed together.	Class Design Document *Prepared by:* class developer(s)	Design review *Reviewed by:* Oversight Team, software manager
Class implementation A class may be stubbed or only partially implemented for a given release.	Prototype code *Prepared by:* class developer(s)	Code review *Reviewed by:* Oversight Team, software manager
Class unit test A class is tested stand-alone with a variety of inputs.	Unit test code *Prepared and tested by:* class developer(s)	Unit test *Verified by:* software manager
	INTEGRATION AND DISTRIBUTION	
Class integration Unit tested class is integrated into an evolving prototype of the ESMF.	ESMF system prototype *Prepared by:* class developer(s), integrator	System test and benchmarking *Verified by:* software manager
User documentation updated Class design documentation is updated and converted to user documentation.	User's Guide *Prepared by:* class developer(s), integrator	Review before software release *Reviewed by:* software manager
System release Code and documentation is released. Defects and requests for features are tracked and incorporated into future releases.	System test *Prepared by:* integrator, software manager	ESMF system release *Evaluated by:* ESS Project, wide user community

2.2 Documents, Reviews and Verification

Documents All documents, including a User's Guide and Reference, will be prepared in a format that easily generates both hardcopy and web-friendly html. We intend to use a documentation generation tool such as ProTeX to automatically create and update portions of our documentation. Most documents will be placed under version control. We will structure design documents so that they can be easily converted to user documentation.

Reviews A variety of reviews will be held, including requirements, design and code reviews. The outcome of a review will either be a pass or, if significant changes are required, the scheduling of another review. The software engineering manager will attend all reviews for coordination.

Verification and benchmarking Class and system tests will be designed so that they verify that the code being tested fulfils its requirements. We plan to develop a synthetic benchmark suite in conjunction with the application groups working on Parts II and III; these benchmarks will be presented in the Software Implementation and Test Plan. The ESS Evaluation Teamand vendors will assist with performance evaluation.

3 Software Tools and Techniques

The following are some of the tools for development and collaboration that we plan to employ:

Configuration management We will likely use CVS for configuration management since it is mature, freely available, and the current community standard. We anticipate maintaining code at the following acceptance levels: *Active* (untested), *Unit tested*, and *Integrated* (code is part of a working ESMF prototype). These levels reflect the completion gates applied to code development shown in Table 1.

Software metrics The software engineering manager will track a simple set of software metrics throughout development in order to evaluate progress towards milestones and predict development schedules. These metrics will include LOSC for each code acceptance level and software engineer, and hours per software engineer spent on the project.

Defect tracking A tool such as Bugzilla will be used to maintain a database of defects and new feature requests.

Collaborative tools We plan to employ weekly teleconferences to keep Oversight Teams in close touch with the Implementation Team, and quarterly face-to-face meetings. We will continue to maintain a mailing list and may explore using web forum software.

4 Source Availability and Distribution

We plan to develop our code in an open source development environment such as SourceForge. In order to meet milestones systematically, we will need to follow a well-defined and carefully managed software engineering plan. Until a complete prototype of the framework is delivered, we do not feel that an open participation model is appropriate. After a prototype is released, we plan to engage the Earth science community in contributing to the ESMF. Community contributions will be integrated into new releases through a maintenance team at NCAR.

Source code and documentation will be distributed via an ESMF website. The website may be an extension of that currently maintained by the CMIWG, as described in Section **. We plan to hold a series of workshops to introduce the broader community to the ESMF.

5 Core ESMF Software Maintenance

NCAR is committed to "institutionalizing" the ESMF, and offering an ongoing program of user support, maintenance (e.g., activities such as porting to new platforms), promotion, and research into improved and extended capabilities after the end of the three-year, NASA-funded activity. We plan to retain some or all members of the Implementation Team at NCAR as core staff to carry out this work.

6 NASA Participation

The software development process described in our suite of proposals does not necessitate the services of an ESMF Integrator or extensive support from NASA software engineering staff. However, we welcome input from NASA technical personnel, and would anticipate detailing NASA involvement during the negotiation process. Please note that the "integrator" described in this section is responsible for assimilating newly developed classes into an ESMF prototype and bears no relation to the ESMF Integrator described in the CAN.

(III) References

References

[1] Airlie Software Council. Principal best practices. Technical report, Software Program Managers Network.

[2] DeLuca, C., Heisey, C.W., Bond, R., and Daly, J.M. A portable object-based parallel library and framework for real-time radar signal processing. In *Scientific Computing in Object-Oriented Parallel Environments*, Berlin, Germany, 1997. Springer-Verlag.

[3] McConnell, S. *Code Complete*. Microsoft Press, 1993.

[4] Office Note Series on Global Modeling and Data Assimilation. Geos-3 data assimilation system. Technical Report DAO Office Note 97-06, NASA Goddard Space Flight Center, July 1997.

[5] Paulk, M.C., Curtis, W., Chrissis, M., and Weber, C.V. Capability maturity model for software, version 1.1. Technical Report CMU/SEI-93-TR-24, DTIC Number ADA263403, Software Engineering Institute, February 1993.

[6] Paulk, M.C., Weber, C.V., Chrissis, M. Garcia, S.M., and Bush, M.W. Key practices of the capability maturity model, version 1.1. Technical Report CMU/SEI-93-TR-25, DTIC Number ADA263432, Software Engineering Institute, February 1993.

[7] Software Engineering Research Laboratory Series. Manager's handbook for software development. Technical Report SEL-84-01, NASA Goddard Space Flight Center, November 1990.

[8] Software Engineering Research Laboratory Series. Recommended approach to software development. Technical Report SEL-81-305, NASA Goddard Space Flight Center, June 1992.

(IV) Biographical Sketches

ARLINDO da SILVA

Data Assimilation Office,
NASA Goddard Space Flight Center
Greenbelt, MD 20771

RESEARCH INTERESTS

My current research interests include techniques for global atmospheric data assimilation, physical-space analysis systems, error covariance modeling, bias estimation and correction, quality control, land-surface, precipitation and aerosol data assimilation, and efficient methods for assimilation of remotely sensed data. Other research interests not in the area of data assimilation include aerosol forcing of climate, hydrological cycle of the subtropics, estimatin of fluxes of heat, momentum and fresh water over the global oceans for observational studies and forcing ocean models.

EDUCATION

B.S.	1982	Physics, Catholic University of Rio de Janeiro, Brazil
M.S.	1984	Physics, Catholic University of Rio de Janeiro, Brazil
Ph.D.	1989	Meteorology, Massachusetts Institute of Technology

EMPLOYMENT

1994–Present	Meteorologist	Data Assimilation Office, NASA Goddard Space Flight Center
1990–1993	Assistant Professor	University of Wisconsin-Milwaukee
1989–1990	Visiting Scientist	Program in Atmospheric and Ocean Sciences Princeton University

RELATED PUBLICATIONS

1. Guo, J., and A. da Silva, 1997: Computational aspects of Goddard's Physical-space Statispical Analysis System (PSAS). In *Numerical simulations in the environmental and earth sciences.*, Garcia et al., Eds., ISBN 052158047, Cambridge University Press, 1997.

2. Dee, D. and A. da Silva, 1998: Data assimilation in the presence of forecast bias. *Q. J. R. Meteor. Soc.*, **124**, 269-295.

3. Joiner, J. and A. da Silva, 1998: Efficient Methods to Assimilate Satellite Retrievals Based on Information Content. *Q. J. R. Met. Soc.*, **124**, 1669–1694.

4. Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, D. Lamich. 1998: Assessing the Effects of Data Selection with the DAO Physical-space Statistical Analysis System. *Mon. Wea. Rev.* . **126**, 2913-26.

5. Dee, D., and A. da Silva, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Wea. Rev.*, **127**, 1822–1834.

6. Sawyer, W., L. Takacs, A. da Silva and P. Lyster, 1999: Parallel grid manipulations in earth science calculations. *Lecture Notes in Computer Science*, **1573**, 666-679. Springer, Berlin.

STEPHEN J. LORD

Environmental Modeling Center
National Centers for Environmental Prediction
NOAA Science Center, Rm. 207
Washington, DC 20233

RESEARCH INTERESTS

My interests are in managing and participating in all aspects of data assimilation and numerical model development for weather and seasonal climate forecasts. As Director of the Environmental Modeling Center, National Centers for Environmental Prediction, I oversee a staff of 90 who are dedicated to improving operational weather, ocean and climate modeling products to support the NWS mission. I have a strong background in data assimilation and tropical meteorology and have done original research on hurricane numerical modeling and data assimilation.

EDUCATION

B.S.	1969	Physics Yale University (cum laude)
M.S.	1975	Atmospheric Sciences University of California at Los Angeles
Ph.D.	1978	Atmospheric Sciences, University of California at Los Angeles

HONORS AND AWARDS

1997	AMS Fellow
1996	NOAA Dept. of Commerce Gold Medal for Implementation of the GFDL Hurricane Model
1993	NOAA Dept. of Commerce Bronze Medal for Applied research on hurricane track prediction

EMPLOYMENT

2000–Present	Director	Environmental Modeling Center, National Centers for Environmental Prediction
1993–2000	Acting Director/Deputy Director	Environmental Modeling Center, National Centers for Environmental Prediction
1989–1993	Meteorologist	National Meteorological Center
1980–1989	Meteorologist	Hurricane Research Division, Atlantic Oceanographic and Meteorological Laboratory

RELATED PUBLICATIONS

1. Pu, Zhao-Xia, S.J. Lord, and E. Kalnay, 1998: Forecast sensitivity with dropsonde data and targeted observations. In press (Tellus)

2. Surgi, N., H.L. Pan, and S.J. Lord, 1998: Improvement of the NCEP global model over the tropics: an evaluation of model performance during the 1995 hurricane season. Mon. Wea. Rev., 126, 1287-1305

3. Tuleya, R. E., and S. J. Lord, 1997: The impact of dropwindsonde data on GFDL hurricane model forecasts using global analyses. Wea. Forecasting, 12, 307-323.

4. Lord, S.J., and J. L. Franklin, 1990: The environment of Hurricane Debby (1982). Part II: Thermydynamic fields. Mon. Wea. Rev., 118, 1444-1459.

5. Lord, S. J., and J. M. Lord, 1998: Vertical velocity structures in an axisymmetric, nonhydrostatic tropical cyclone model. J. Atmos. Sci., 45, 1453-1461.

MICHELE RIENECKER

Oceans and Ice Branch
NASA/Goddard Space Flight Center
Greenbelt, MD 20771

RESEARCH INTERESTS

My research interests include various aspects to understand the ocean's role in climate variability. Within NSIPP, I use numerical ocean models and develop ocean data assimilation methods for prediction of El Niño using coupled general circulation models. I am also interested in numerical analysis and in parallel computing algorithms.

EDUCATION

B.S.	1974	First Class Honours in Applied Mathematics Univeristy of Queensland, Australia
Ph.D.	1980	Applied Mathematics University of Adelaide, Australia

EMPLOYMENT

1991–Present	Oceanographer	NASA/Goddard Space Flight Center
1989–1991	Research Scientist	USRA, NASA/GSFC
1988–1989	NRC Senior RRA	NASA/GSFC
1986–1988	Scientist II	Institute for Naval Oceanography
1982–1986	Adjunct Research Professor	Naval Postgraduate School, Monterey

RELATED PUBLICATIONS

1. Borovikov, A., M.M. Rienecker and P.S. Schopf (2000), Surface heat balance in the equatorial ocean: climatology and the warming event of 1994-95, J. Clim, (in press).

2. Keppenne, C.L. and M.M. Rienecker (1999), Assimilation of temperature data into an ocean general circulation model with a parallel Ensemble Kalman Filter. *Proceedings of 3rd WMO Symposium on Data Assimilation in Meteorology and Oceanography*, Quebec.

3. Rienecker, M.M., A. Borovikov, C. Keppenne, and D. Adamec (1999), Impact of multivariate assimilation on estimates of the state of the tropical Pacific Ocean. *Proceedings of 3rd WMO Symposium on Data Assimilation in Meteorology and Oceanography*, Quebec.

4. Rienecker, M.M. and D. Adamec (1995), Assimilation of altimeter data into a quasigeostrophic model using optimal interpolation and eofs. J. Marine Systems, 6, 125–143.

CHRISTIAN L. KEPPENNE

Code 971
NASA/Goddard Space Flight Center
Greenbelt, Maryland 20771

RESEARCH INTERESTS

My research interests include scientific computing, parallel applications development, geophysical fluid dynamics, and data assimilation.

EDUCATION

B.S.	1986	Highest Honors, Civil Engineering Catholic University of Louvain, Belgium
M.S.	1987	Highest Honors, Applied Science Catholic University of Louvain, Belgium
Ph.D.	1989	Highest Honors, Applied Science Catholic University of Louvain, Belgium

EMPLOYMENT

1998–Present	Senior Research Scientist	General Sciences Corporation
1990–1997	Research Scientist	JPL, California Institute of Technology
1988–1990	Research Scientist	Dept. of Atmospheric Sciences, UCLA
1986–1988	Research Fellow	Belgian National Foundation for Scientific Research

RELATED PUBLICATIONS

1. Keppenne C.L. (2000), Data assimilation into a primitive equation model using a parallel ensemble Kalman filter, *Mon. Wea. Rev.*, 128, 1971-1981.

2. Keppenne C.L., S.L. Marcus, M. Kimoto, and M. Ghil (2000), Intraseasonal variability in a two-layer model and observations, *J. Atmos. Sci.*, 57, 1010-1028.

3. Keppenne C.L. and M.M.Rienecker (1999), Massively parallel sequential assimilation of temperature data from the TAO array into an ocean general circulation model, *Proceedings, 3rd World Meteorological Organization Symposium on the Assimilation of Observations in Meteorology and Oceanography*, held at Quebec, Canada, June 7-11, 1999.

4. Banfield D.J., A.P. Ingersoll, and C.L. Keppenne (1995), A steady-state Kalman filter for assimilating data from a single polar-orbiting satellite, *J. Atmos. Sci.*, 52, 738-753.

5. Keppenne C.L., and M. Ghil (1993), Adaptive filtering and prediction of noisy multi-variate signals: an application to atmospheric angular momentum, *Intl. J. Bifurcations and Chaos*, 3, 625-634.

6. Keppenne C.L., and M. Ghil (1992), Adaptive Spectral Analysis and Prediction of the Southern Oscillation Index, *J. Geophys. Res.*, 97, 20449-20554.

PETER M. LYSTER

University of Maryland Earth System Science Interdisciplinary Center and
NASA Goddard Space Flight Center Data Assimilation Office

RESEARCH INTERESTS

I work on atmospheric data assimilation and space weather, focusing on new algorithms, the scientific software development process, and high-performance computing.

EDUCATION

B.S.Hons. Physics 1976. B.Eng.Hons(EE). 1977. The University of Adelaide, Australia
M.S. Num. Anal. 1981. Ph.D. Plasma Physics 1987. Cornell University

EMPLOYMENT

1994–Present	Research Scientist, University of Maryland
1990–1993	Member of Technical Staff, Jet Propulsion Laboratory
1986–1990	Postdoctoral Research Fellow, The Institute for Fusion Studies University of Texas

RELATED PUBLICATIONS

1. P. M. Lyster: Final Report NASA Grand Challenge Applications and Enabling Scalable Computing Testbed in Support of High Performance Computing: Four Dimensional Data Assimilation, (July 4, 2000). Available at `http://dao.gsfc.nasa.gov/DAO_people/lys/hpccfinal`

2. R. Ménard, S. E. Cohn, L.-P. Chang, and P. M. Lyster: Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part I: Formulation. *Mon. Wea. Rev.*, **128**, 2654-2671 (2000).

3. W. Sawyer, R. Lucchesi, P. M. Lyster, L. L Takacs, A. Molod, J. Larson, S. Nebuda, C. Pabon-Ortiz: Parallelization of DAO Atmospheric General Circulation Model. *Proc. Fourth International Workshop on Applied Parallel Computing (PARA98)*, Springer-Verlag, Lecture Notes in Computer Science, Vol. 1541, pp. 510-514, ISBN 3-540-65414-3 (1998).

4. P. M. Lyster, J. W. Larson, J. Guo, W. Sawyer, A. da Silva, and I. Štajner: Progress in the Parallel Implementation of the Physical-space Statistical Analysis System (PSAS). *Making it's Mark: Proc. Seventh ECMWF Workshop on the Use of Parallel Processors in Meteorology*, Eds. G-R. Hoffmann and N. Kreitz, 382-393 (World Scientific, 504pp, 1998). Available at `http://www.wspc.com/books/compsci/3679.html`

5. P. M. Lyster, J. W. Larson, W. Sawyer, C. H. Q. Ding, J. Guo, A. M. da Silva, L. L. Takacs: Parallel Computing at NASA Data Assimilation Office (DAO). *Proc. Supercomputing97*, San Jose, November 1997. Available at `http://www.supercomp.org/sc97/proceedings/TECH/LYSTER/`

6. P. M. Lyster, S. E. Cohn, R. Ménard, L.-P. Chang, S.-J. Lin, and R. Olsen: Parallel Implementation of a Kalman Filter for Constituent Data Assimilation. *Mon. Wea. Rev.*, **125**, 1674-1686 (1997).

7. P. M. Lyster, P. C. Liewer, R. D. Ferraro, and V. K. Decyk: Implementation and Characterization of Three-Dimensional Particle-In-Cell Codes on Multiple-Instruction-Multiple-Data Parallel Supercomputers. *Comp. Phys.*, **9(4)**, 420-432 (1995).

BYRON A. BOVILLE

National Center for Atmospheric Research,
P.O. Box 3000, Boulder CO 80307

RESEARCH INTERESTS

My research has concentrated on developing and applying general circulation models of the lower and middle atmosphere for studies of atmospheric dynamics and climate. I have been one of the central figures in both the scientific and computational development of 4 generations of the NCAR atmospheric general circulation model. More recently, I have concentrated on coupled ocean-atmosphere modeling and was co-chair of the team which developed the NCAR Climate System Model (CSM). I am currently interested in the climate impact of solar variability and the role of the middle atmosphere in climate variability and climate change.

EDUCATION

B.S.	1975	1st Class Honours in Meteorology, McGill University, Montreal, Canada
Ph.D.	1979	Atmospheric Sciences, University of Washington, Seattle, Washington

EMPLOYMENT

1999–Present	Head	Climate Modeling Section, Climate and Global Dynamics Division National Center for Atmospheric Research
1992–Present	Senior Scientist	National Center for Atmospheric Research
1981–1992	Scientist I-III	National Center for Atmospheric Research
1979–1981	Postdoc	Advanced Study Program National Center for Atmospheric Research

RELATED PUBLICATIONS

1. Boville, B. A., J. T. Kiehl, P. J. Rasch, and F. O. Bryan, 2001: Improvements to the NCAR CSM-1 for transient climate simulations. J. Climate, 14, in press.

2. Boville, B. A., 2000: Toward a complete model of the climate system. In "Numerical modeling of the global atmosphere in the climate system", P. Mote and A. O'Neill, eds., Kluwer Academic Publishers, 419–442.

3. Kiehl, J. T., J. J. Hack, G. B. Bonan, B. A. Boville, D. L. Williamson, P. J. Rasch, 1998: The National Center for Atmospheric Research Community Climate Model: CCM3. J. Climate, 9, 1131–1149.

4. Boville, B. A., and P. R. Gent, 1998: The NCAR climate system model, version one. J. Climate, 11, 1115–1130.

5. Hack, J. J., J. M. Rosinski, D. L. Williamson, B. A. Boville, and J. E. Truesdale, 1995: Computational design of the NCAR community climate model. Parallel Computing, 21, 1545–1569.

CECELIA DeLUCA

National Center for Atmospheric Research,
P.O. Box 3000, Boulder CO 80307

RESEARCH INTERESTS

My interests include the design of large, high-performance software systems, particularly those relating to atmospheric science; parallel algorithms; real-time systems, and software engineering tools and processes. I was a design lead on the development of the Space-Time Adaptive Processing Library (STAPL) parallel framework for real-time radar applications. STAPL is an integral part of multiple operational next-generation radar systems and has been ported to several platforms. It extends the serial Vector Signal and Image Processing Library (VSIPL) standard to SMP-cluster architectures. Previous projects have included the development of parallel codes for the simulation of middle atmospheric dynamics, atmospheric chemistry, and remote sensing of atmospheric temperatures.

EDUCATION

A.L.B.	1990	Liberal Arts/Social Sciences, Harvard University, Cambridge, MA
M.S.	1994	General Engineering, Boston University, Boston, MA
M.S.	1996	Meteorology, Massachusetts Institute of Technology, Cambridge, MA

AWARDS

1994 Boston University College of Engineering Outstanding Achievement Award, first in graduating class

EMPLOYMENT

1999–Present	Software Engineer	Scientific Computing Division, National Center for Atmospheric Research
1998–1999	Lead Software Engineer	MIT Lincoln Laboratory
1996–1998	Software Engineer	MIT Lincoln Laboratory
1993–1994	Manager, Technical Support	Omnet Communications

RELATED PUBLICATIONS

1. Dickinson, R.E., S.E. Zebiak, J.L. Anderson, M. Blackmon, C. DeLuca, T. Hogan, M. Iredell, M. Ji, R. Rood, M. Suarez, K. Taylor, "Need for Infrastructure and Commonality in Climate and Weather Prediction Codes and Data," submitted to *Bulletin of the American Meteorological Society*, 2000.

2. DeLuca, C., C. Heisey, R. Bond and J. Daly, " A Portable, Object-Based Parallel Library and Layered Framework for Real-Time Radar Signal Processing," In *Proceedings of Scientific Computing in Object-Oriented Parallel Environments*, ISCOPE 1997.

3. Heisey, C., C. Adamo, M. Arakawa, P. Baggeroer, J. Daly, C. DeLuca, W. Dale Hall, K. Pickard, and H. A. Spang, "Implementation of the STAP Library and Framework (STAPL) for Real-Time Matrix-Based Signal Processing," In *Abstracts of High Performance Embedded Computing*, HPEC 1998.

JOHN MARSHALL

Department of Earth, Atmospheric and Planetary Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139

RESEARCH INTERESTS

My research is directed at understanding key components of the general circulation of the atmosphere and ocean and the development of models to study them. I am interested in a variety of problems in geophysical fluid dynamics and their role in climate, ranging from rotating convection, the global circulation of the ocean and air-sea interaction. I use and develop numerical models of the atmosphere, ocean and climate.

EDUCATION

B.S.	1976	First Class Honors in Physics, Imperial College, London
Ph.D.	1980	Physics Imperial College

EMPLOYMENT

1992–Present	Professor	Massachusetts Institute of Technology
1992	Associate Professor	Massachusetts Institute of Technology
1991–1992	Reader in Physics	Imperial College
1984–1990	Lecturer in Physics	Imperial College
1982–1983	Post-doctoral fellow	University of Oxford

RELATED PUBLICATIONS

1. Marshall, J., C. Hill, L. Perelman, and A. Adcroft, (1997) Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling, J. Geophysical Res., 102(C3), 5733-5752.

2. Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey, (1997) A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers, J. Geophysical Res., 102(C3), 5753-5766.

3. Adcroft, A.J., Hill, C.N. and J. Marshall, (1997) Representation of topography by shaved cells in a height coordinate ocean model Mon Wea Rev, vol 125, 2293-2315

4. Marshall, J., Jones, H. and C. Hill, (1998) Efficient ocean modeling using non-hydrostatic algorithms Journal of Marine Systems, 18, 115-134

5. Shaw; A. Arvind, Cho, K.-C., Hill, C., Johnson, R.P. and Marshall, J. (1998) A comparison of implicitly parallel multi-threaded and data-parallel implementations of an ocean model based on the Navier-Stokes equations. J. of Parallel and Distributed Computing, vol 48, No 1, 1-51

(V) Milestones, Schedule and Costs

The dates on these milestones assume that the first payment is received June 2001.

<table>
<tr><th>No.</th><th>Label</th><th>Milestone Title</th><th>Expected Completion Date</th><th>Advance Payment Amount</th></tr>
<tr><td>1</td><td>A</td><td>Software engineering plan completed.
Deliver draft Developer's Guide specifying software procedures and conventions.</td><td>Sept. 2001</td><td>$124,884</td></tr>
<tr><td>2</td><td>E</td><td>Code baseline completed.
Baseline atmospheric and oceanic data assimilation systems delivered on ESS testbed (except for NCEP system, see below):
DAO: OpenMP version of DAO's global 3D-VAR system including the Finite-volume Climate Community Model Version 3 (fvCCM3) and the Physical-space Statistical Analysis System (PSAS). The FvCCM will run at 2x2.5L55 resolution (2x2.5° horizontal, 55 vertical layers); PSAS will run at 2x2.5L25 resolution with ≈200K observations/day.
NCEP: The NCEP global atmospheric forecast and analysis codes will be baselined on NCEP's home system. The baseline will indicate which platforms the codes run on, how many different configurations of processors and nodes the codes can run under, and how well the codes perform on these platforms and under these configurations. The baseline codes will be the current operational system which runs at T170 resolution with 2 tracers using 250K observations per day.
NSIPP: GEMS version of NSIPP's global OI system including V4 of the Poseidon Ocean model and the PSAS-like OI analysis system. Poseidon and the ODAS will run at $\frac{2}{3} \times 1.25$L20 resolution with ≈10K observations/day.
MIT: The 'wrapper layer' version of the adjoint operators currently employed will be bench-marked for scaling and portability. Resolution is 2×2L22 with 10 iterations.
Provide code scaling. The metric will be throughput (number of simulated days per wallclock day). Documented source code will be delivered via the Web.</td><td>Nov. 2001</td><td>$291,397</td></tr>
</table>

(continued)

No.	Label	Milestone Title	Expected Completion Date	Advance Payment Amount
3	H	*Design policy for interoperability and community delivery agreed on.* Deliver Architecture Document, Implementation Report, Draft Software Implementation and Test Plan. Prototype code used in the Implementation Study will be delivered via the Web. Identify a broader community interested in reviewing/testing the ESMF and mechanism for feedback.	Dec. 2001	$416,281
4	B	*First Annual Report delivered.* Submit FY01 Annual Report via the Web.	June 2002	$125,616
5	F	*First code improvement completed.* Improved versions of at least 2 of the following data assimilation systems: • **DAO**: MPI versions of model (fvCCM4) and analysis (mpiPSAS). The FvCCM will run at 1x1.25L55 resolution (twice the resolution of baseline) while PSAS will run at 2x2.5L25 resolution with ≈300K observations/day with at least the same throughput as the baseline. • **NCEP:** Using the prototype ESMF, NCEP will port both its global atmospheric forecast and analysis codes to at least one new platform. The forecast and analysis codes will interoperate with each other using the frameworks classes. The performance of the codes should not be more than 5% worse than the baseline code using the same computer resources. • **NSIPP:** GEMS versions of OGCM and ODAS, with ODAS upgraded to include the Ensemble Kalman Filter with 100 ensemble members at baseline resolution. The OGCM with OI will run at baseline resolution with double the throughput. • **MIT:** The problem size for benchmark cases will be increased by a factor of two in spatial and temporal resolution. Provide code scaling and deliver documented source code over the web.	July 2002	$293,104

(continued)

No.	Label	Milestone Title	Expected Comple-tion Date	Advance Payment Amount
6	I	*Interoperability prototype from Milestone "H" tested with improved codes.* DAO/NCEP atmospheric data assimilation systems will utilize ESMF services to access and distribute common observational data base. NSIPP OI and EnKF ODAS will utilize ESMF services to access and distribute ocean observational data base on GODAE server. DAO/NCEP/NSIPP models will utilize ESMF classes/methods for interfacing with analyses. The mitGCM adjoint sensitivity test cases will be adapted to employ framework primitives. Documented source code delivered via the Web.	Nov. 2002	$418,720
7	C	*Second Annual report delivered.* Submit FY02 Annual Report via the Web.	June 2003	$106,384

(continued)

No.	Label	Milestone Title	Expected Comple-tion Date	Advance Payment Amount
8	G	*Second code improvement completed.* Improved versions of data assimilation systems. At least 3 of the following: • **DAO**: The FvCCM will run at 0.5x0.625L55 resolution while PSAS will run at 1x1.25L25 resolution and ≈400K observations/day with at least the same troughput as the baseline. • **NCEP:** Using the ESMF, NCEP will demonstrate that its global atmospheric forecast and analysis codes can interoperate with codes from other centers. Furthermore, the performance of these codes using the ESMF will be sufficiently efficient to be usable as the basis for a community version of NCEP's data assimilation system. The performance of the codes should not be more than 5% worse than the baseline code using the same computer resources. Also at this time, by using more computer resources, the full NCEP analysis and forecast system will be run at T254 resolution with 3 tracers using 500K observations per day in the same time window as the baseline system. • **NSIPP:** Efficient implementation using ESMF to allow doubled problem: with twice as many grid points or double the ensemble size. Use frameworks for ODAS in a coupled model configuration with the NSIPP AGCM from part II. • **MIT:** Increased problem size and improved scalability to allow more in-core time-history storage. Demonstrate adjoint test case scaling efficiently on affordable cluster hardware. Provide code scaling and deliver documented source code over the web.	July 2003	$294,897

(continued)

<table>
<tr><th>No.</th><th>Label</th><th>Milestone Title</th><th>Expected Completion Date</th><th>Advance Payment Amount</th></tr>
<tr><td>9</td><td>J</td><td>Full interoperability demonstrated using improved codes.
Implementation of interoperable atmospheric data assimilation applications:
DAO model with DAO analysis
NCEP model with DAO analysis
NCEP model with NCEP analysis
DAO model with NCEP analysis
Implementation of interoperable ocean data assimilation applications by at least one of:
NSIPP OI and EnKF with common ESMF-compliant analysis code, external to the covariance calculation
NSIPP ODAS with DAO global solvers.
Demonstrate mitGCM adjoint sensitivity test case operating with the ECCO optimisation environment.

ESMF will be used to couple model/analysis components as well as to access observational database, with partial utilization of ESMF services internally within models/analyses.</td><td>Nov. 2003</td><td>$421,281</td></tr>
<tr><td>10</td><td>K</td><td>Customer delivery accomplished.
Port the framework to a new platform within 1 month and demonstrate its operation. Deliver a user-friendly website where the ESMF source code and User's Guide can be obtained, and hold a workshop on framework usage.</td><td>May 2004</td><td>$15,000</td></tr>
<tr><td>11</td><td>D</td><td>Final Report delivered.
Submit Final Report to ESS via the Web.</td><td>June 2004</td><td>$5,000</td></tr>
</table>

(VI) Education and Public Outreach

Education and Public Outreach Addendum for the Earth System Modeling Framework

Background

Enhancing the science, math, and technology literacy of our population is a vital national interest, as has been expressed eloquently in the national planning document, *Science in the National Interest*:

> "The scientifically literate society that America will need to face in the 21st century will require orientation to science early in life and frequent reinforcement. The lifelong responsibilities of citizenship increasingly rely on scientific and technological literacy for informed choices. Our scientific community must contribute more strongly to broad public understanding and appreciation of science. Our education system must provide the necessary intellectual tools at twenty-first century standards."
>
> *Science in the National Interest*

This statement is particularly relevant for the geosciences. Recent test results have shown that middle and high school students have only a poor understanding of key concepts that could otherwise help them understand the world around them. Improvements in geoscience education have been identified as a primary area for focusing efforts to improve science literacy of children. Environmental concerns are becoming increasingly pressing and complex, requiring citizens to have a much better understanding of Earth systems and processes, and the methods scientists use to unravel this complexity to develop an understanding of the underlying systems.

It is unfortunate that there are few K-12 educators with direct training in the geosciences, but this may be reflected in the low content mastery displayed by students in this subject area nationwide. In order to assist students in building the knowledge they need to be competent in the geosciences, training and educational opportunities need to be made available that will provide educators with the background they need to be confident and enthusiastic models of interest and knowledge in the relevant subject areas and skills sets. Similarly, educators need access to resources that are readily useable in their classrooms, are integrated into the curriculum requirements within which the educators operate, and are relevant and engaging to both students and teachers.

E/PO Objectives and Planned Activities

The objective of the education and public outreach addendum of the suite of Earth System Modeling Framework proposals is to develop and implement professional development opportunities on the Earth system sciences for middle and high school educators. Building upon and extending existing resources, a series of nationally-advertised professional development workshops for middle school and high school educators will utilize the concept of global change to explore aspects of the coupled Earth system. This project will be implemented by UCAR scientists, education specialists, and nationally recognized experts in professional development in the geosciences for middle and high school educators.

Workshops

The most effective way to improve the scientific literacy of students is to work to improve the training of those responsible for educating them. The focus of this effort will be to build upon our extensive experience and existing resources to develop and implement a series of professional development opportunities for middle and high school educators that provides the training they need to enrich the learning of their students. In order to ensure the highest leverage to educators, we propose to target our training activities on educators with a proven track-record of providing training opportunities in their own regions or districts.

The workshops proposed here will extend upon the highly successful model used for the Michigan Space Grant Consortium (MSGC) Geoscience Education Workshop (see below). The yearly workshop series will target 20 educators per year, and will include an intensive 2-week summer semester at NCAR in Boulder, followed by 2 additional 2-day professional development opportunities during the school year. In addition to providing background content on the Earth system that highlights issues of global change and climate change, the workshops will provide opportunities to explore modeling concepts through state-of-the-art Earth system models (the science focus of this suite of proposals) as well as simplified STELLA models of components of the Earth system relevant to curriculum needs of educators as expressed in the National Science Education Standards. Educators will be trained on easy to implement hands-on, inquiry-based classroom activities that illustrate concepts of modeling and aspects of the Earth system. The workshop will include opportunities for educators to gain field experience - for observations and data collection that will be utilized in their studies and also expand their appreciation of the Earth sciences. Finally, the workshop will provide guidance on techniques participants can use to bring the training they have received back to educators in their own districts.

The workshop opportunities will be advertised nationally. Participants will be selected from the applicant pool based on their previous experience and success in regional or local professional development for science educators, as well as our programmatic desire to have a diverse and geographically distributed set of workshop participants. Travel and living expenses for workshop participants will be covered during the workshops, and each participant will receive a stipend to support their efforts to bring the training they have been provided back to educators in their own districts. In return for the training and support they have received, participants will be expected to train an additional 20 educators per year in their own districts. Materials needed for this training will be provided from the project.

A dedicated project web site will be developed that provides access to information about the project, all lectures, labs, and other training resources, and also provides a venue for communication among project participants.

Previous Relevant Education and Outreach Activities of Project

Personnel Science and education personnel participating in this proposal have extensive experience in development of highly-valued educational resources and successful professional development opportunities for K-12 educators. Over the past decade, UCAR has developed a number of professional development resources with funding from NSF for middle and high school educators on global change, climate change, and the ways in which components of the Earth system are interconnected. In addition, NCAR has been responsible for the development of an influential series of documents *Reports to the Nation* that highlight our understanding of the Earth system and the potential contributions of humans in changing our climate. With funding from NSF, UCAR developed Project LEARN, in which science education faculty, scientists, and professional development specialists provided training to over 2000 middle school educators on the atmospheric sciences. After six years of funding, resources developed through this program are being made available on-line to support educators nationwide.

Prior to moving to UCAR, Dr. Roberta Johnson (the Director of the UCAR Office for Education and Outreach) developed, in collaboration with colleagues, a highly successful workshop for middle and high school Earth science educators while she was the Director of the MSGC (funded by NASA).

This workshop utilized the concept of global change as a vehicle for exploring the breadth of the geosciences, covering topics ranging from recycling to anthropogenic effects on ozone and greenhouse gases to the role of plate tectonics in determining planetary climates. Workshop activities included a combination of lecture, computer-based training and activities (with an emphasis on modeling illustrated through STELLA), training on hands-on inquiry-based activities, field experience, and project development. Resources utilized in the workshop included content and activities available through the NASA sponsored Windows to the Universe project (also PI-d by Dr. Johnson), content and labs available through the NASA and NSF sponsored Introduction to Global Change Course developed by Dr. Timothy Killeen (the PI of this suite of proposals) while he was faculty at the University of Michigan, as well as other quality resources developed by other groups addressing relevant geoscience topics (Topex-Poseidon CD, Ozone Challenge, Blue-Skies, etc).

Management and Coordination

Dr. Johnson will be responsible for the directing the educational activities described in this addendum. She will be assisted by the staff of the UCAR Office of Education and Outreach, including Ms. Susan Foster, who will manage the project and assist in design and implementation of the workshop. Dr. Killeen and other selected science Co-Is on the project will participate in the educational effort by giving lectures to project participants and participating in other workshop activities. Mr. David Mastie and Mr. Parker Pennington IV, both recognized experts in teacher training in the Earth sciences at the middle and high school levels, will collaborate in the effort and continue in the roles they have had in the MSGC Geoscience Education Workshop as lead teacher trainers, working with local UCAR EO staff.